CLASSIFICATION OF MENTAL TASKS USING DE-NOISED EEG SIGNALS

Mohd Daud, S.[1] (IEEE Student Member) Yunus.J.[2]
Universiti Teknologi Malaysia, 54100, Kuala Lumpur, Malaysia
[1]salwani@citycampus.utm.my [2]jasmy@fke.utm.my

Abstract: The wavelet based de-noising can be employed with the combination of different kind of threshold parameters, threshold operators, mother wavelets and threshold rescaling methods. The central issue in wavelet based de-noising method is the selection of an appropriate threshold parameters. If the threshold is too small, the signal is still noisy but if it is too large, important signal features might lost. This study will investigate the effectiveness of four types of threshold parameters i.e. threshold selections based on Stein's Unbiased Risk Estimate (SURE), Universal, Heuristic and Minimax. Autoregressive Burg model with order six is employed to extract relevant features from the clean signals. These features are classified into five classes of mental tasks via an artificial neural network. The results show that the rate of correct classification varies with different thresholds. From this study, it shows that the de-noised EEG signal with heuristic threshold selection outperforms the others. Soft thresholding procedure and sym8 as the mother wavelet are adopted in this study.

Keywords : EEG, de-noised, wavelet shrinkage, threshold

1. Introduction

Electroencephalogram (EEG) is a non-stationary and noisy signal recorded non-invasively via scalp electrodes. Classification of the mental tasks can achieve higher accuracy if noise can be suppressed effectively. Wavelet transform (WT) has revolutionized signal and image processing over the past two decades. An important part of signal processing is to eliminate noise or de-noising i.e. recovering the 'true' signal from the noisy data. Wavelet had performed effectively in this field. Donoho and Johnstone principally developed de-noising by threshold in the wavelet domain [1-2].

Autoregressive (AR) Burg model with order six is employed to extract relevant features from the clean signal. These features are then classified into five mental tasks of interest i.e. baseline, multiplication, letter composing, rotation of a 3-D figure and visual counting. The classifier used is a simple multilayer feed forward back-propagation neural network. These mental tasks classification are useful for brain-computer interface system specially developed for severe physical disabilities individuals.

0-7803-8406-7/04/$20.00 © 2004 IEEE.

2. Wavelet-based De-noising

In wavelet analysis, linear combination of wavelet functions consisting of mother wavelet function, $\psi(t)$ and scaling function, $\phi(t)$ are used to represent a signal, y(t) as follows :

$$y(t) \approx \sum_k a_{j,k}\phi_{j,k}(t) + \sum_k d_{j,k}\psi_{j,k}(t) + \ldots + \sum_k d_{1,k}\psi_{1,k}(t) \quad (1)$$

where j is the number of multi-resolution levels and k ranges from 1 to the number of coefficients in the specified component. The set of coefficients $a_{j,k}$, $d_{j,k}, \ldots, d_{1,k}$ is the wavelet transform of the original signal. The coefficients at level d_j represent signal at lower frequency band than the coefficients at level d_{j-1}. The coefficients of a_j represent an average of the original signal.

Donoho and Johnstone proposes an algorithm to suppress noise in a signal known as wavelet de-noising [1-2]. Suppose a signal in additive white Gaussian noise is represented by:

$$y(i) = f(t_i) + z(i) \quad \text{for } i = 1,\ldots,N. \quad (2)$$

y(i) represents the noisy signal, $f(t_i)$ is the deterministic true signal, the Gaussian white noise with independent identical distribution (i.i.d.), z(i) modeled with mean zero and known variance, σ^2. The goal of de-noising is to recover f by optimizing the mean-squared error (MSE) [1] defined as follows:

$$\frac{1}{N}E\left\|\hat{f} - f\right\|^2 = \frac{1}{N}\sum_{i=1}^{N} E\left(\hat{f}(i/N) - f(i/N)\right)^2 \quad (3)$$

where $\hat{f}$ is the estimator of f. We use soft threshold method to eliminate noise from the wavelet coefficients by replacing the coefficients that are in the range $[-\delta,+\delta]$ with zero while other coefficients are being reduced by a threshold value. Soft thresholding has nice mathematical properties and does not create discontinuities [1]. The soft threshold function is :

$$D(x,\delta) = \begin{cases} 0 & for\,|x| \le \delta \\ sign(x)(|x - \delta|) & for\,|x| > \delta \end{cases} \quad (4)$$

The second part of the equation shows that the coefficients are shrunken by threshold value, δ when they are above the threshold parameter. The three steps [1-2] in wavelet shrinkage de-noising procedure are as follows :

(1). Apply WT with J levels to the signal.
(2). Apply the non-linearly soft threshold function to the wavelet coefficients. Then the estimate coefficients are obtained based on the selected threshold rule.
(3). Use inverse WT on the shrunken wavelet coefficients.

The main problem in de-noising procedure is to choose an appropriate threshold parameter since the signal obtained will still has the noisy components if the value is too small [3]. On the other hand, a large threshold will remove important signal features. Four types of threshold parameters studied in this paper are:
(1). Stein's Unbiased Risk Estimate (SURE): Charles Stein has developed a method for estimating the loss $\left\|\hat{f} - f\right\|^2$ in an unbiased way [2]. Threshold selection based on Stein's Unbiased Risk Estimate (SURE) will select a near optimal threshold at a resolution level according to:

$$\delta = \arg\min_{\delta \geq 0} SURE(\delta; x) \tag{5}$$

and the unbiased estimate of risk is:

$$SURE(\delta; x) = N - 2.\#\{i : |x_i| \leq \delta\} + \sum_{i=1}^{N} (|x_i| \wedge \delta)^2 \tag{6}$$

where N is the number of wavelet coefficients, x [2], [4].
(2). Universal threshold uses a fixed form threshold and is defined as:

$$\delta = \sqrt{2\log(N)} \tag{7}$$

where δ is the threshold value and N is the length of data samples.
(3). Heuristic version of threshold selection which combines the two previous options. The SURE threshold does not perform well when the wavelet coefficients are extremely sparse [2]. A test for the sparseness [2], [4] is as follows:

$$sparsity = \frac{1}{N}\left(\sum_{i=1}^{N} |x_i|^2 - N\right) \tag{8}$$

$$critical = \sqrt{\frac{1}{N}\left(\frac{\log N}{\log 2}\right)^3} \tag{9}$$

If $sparsity \leq critical$, universal threshold is used otherwise threshold selection based on SURE is adopted. If SURE is used in situations of extreme sparseness, the SURE estimates will be very noisy.
(4). Minimax is based on minimax principle uses a fixed threshold to yield the minimax mean square error (MSE) [4], [5] that is obtained for the worst function in a given set, when compared against an ideal procedure. The threshold selection [5] is given by:

$$\delta = 0.3936 + 0.1829 \times \frac{\log(N)}{\log(2)} \tag{10}$$

It is possible to rescale or adapt threshold parameters according to multi-resolution levels either to median value of the detail coefficients of the first level or to every detail coefficients at every level [2],[4]. Many different kinds of wavelet shrinkage de-noising procedures can be generated by combining different choices for wavelet function (also known as mother wavelet), thresholding rules (δ) and operators (either hard or soft threshold) [3]. In this study, we compare the effectiveness of each threshold parameters to de-noise EEG signals using soft thresholding procedure with sym8 as the mother wavelet.

3. Feature Extraction Using AR Burg

Keirn and Aunon [6] had shown that mental task feature extraction using AR Burg model of order 6 outperformed Burg spectrum method and Wiener-Khinchine method. The EEG signals are modeled by zero-mean, stationary and non-deterministic with AR process of order p is given by :

$$x(k) = -\sum_{l=1}^{p} a_l^n x(k-l) \quad + \quad e(k) \tag{11}$$

where p is the model order, x(k) is the data of the signal at sampled point k, a_l^n are the AR coefficients and e(k) represents the prediction error of the signal x. We extract the de-noised signal using AR Burg model of order 6 in this study.

4. Classification Using Artificial Neural Network

Artificial neural network (ANN) is a set of connected input/output units (neurons) where each connection has a weight associated with it. Among the advantages of neural networks include their robustness to noisy data, output may be discrete, real-valued or combination of both and high prediction accuracy. One example of network architecture is multi-layer feed-forward neural network as shown in Figure 1. Researchers [7] had pointed out that classification of EEG features with neural networks yields better classification accuracy compare with other linear methods.

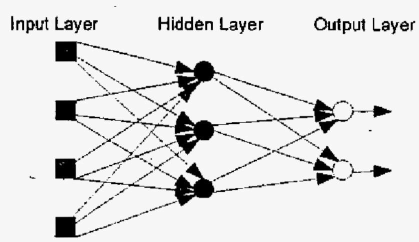

Figure 1 Multi-layer Feed-Forward Neural Network

This NN is being trained by the back-propagation algorithm in a supervised manner. Basically, the back-propagation algorithm is based on the error-correcting learning rule. It consists of two passes through different layers of the network : a forward pass and backward pass.

Details about this NN can be read from the relevant sources.

5. Experimental Results

The data used in this study is taken from the works of previous researchers [6] and were collected using the following procedure. Subjects were selected and placed in a dim, sound controlled room with scalp electrodes at positions C3, C4, P3, P4 O1 and O2 and referenced to two electrically linked mastoids at A1 and A2. Data were sampled at 250 Hz and the electrodes were connected via a bank of amplifiers with analog band pass filters from 0.1 to 100 Hz. Every subject was asked to perform five mental tasks i.e. relaxing or resting (think nothing) with eyes closed, mentally solving multiplication problem, mentally composing a letter to a friend, visualize a sequence of numbers being written on a blackboard and rotate a three dimensional block. Each task was recorded for 10 s and every subject performed each task for five trials.

The de-noised EEG data from each channel was divided into half second segment and overlapped by a quarter second segment which produces 32 segments. AR Burg model with order 6 was employed for each channel independently for these data giving 6 coefficients per channel and a total of 36 coefficients for each task were obtained. These AR features were classified using neural network with 36 inputs, 20 hidden units and five output units. This NN was trained by back-propagation with learning rate, $\eta = 0.1$ and training will stopped after 2000 iterations or when it is validated. The training data was selected from the full set of five trials from a subject: one trial was selected as test data, another one was for validation set and the remaining three trials were compiled into one set of training data.

Figure 2 shows the waveforms of noisy and de-noised signals from a channel with different thresholds using default median value. The performance of the classifier based on the test data is given in Tables 1-3 for different thresholds used.

Table 1: Percentage of test data correctly classified for each task using default median value (=1)

Task Thresh.old	Rest	Multi -ply	Let-ter	Rota-te	Count	Ave. across tasks
Heuristic	86.3	73.8	85.6	79.4	86.3	82.3
Universal	76.9	68.8	83.8	62.5	60.0	70.4
SURE	80.6	81.3	70.6	76.3	68.1	75.4
Minimax	83.8	74.4	73.8	68.1	71.9	74.4

Table 2: Percentage of test data correctly classified for each task using single* median value

Task Thresh.old	Rest	Multi -ply	Let-ter	Rota-te	Count	Ave. across tasks
Heuristic	61.3	76.9	58.8	63.1	63.1	64.6
Universal	56.9	54.4	50.0	58.8	67.5	52.5
SURE	44.4	40.0	48.8	60.0	69.4	52.5
Minimax	54.4	56.9	50.0	58.1	49.4	53.8

Table 3: Percentage of test data correctly classified for each task using adapt* median value

Task Thresh.old	Rest	Multi -ply	Letter	Rota-te	Count	Ave. across tasks
Heuristic	70.0	42.5	45.0	69.4	55.0	56.4
Universal	58.1	61.3	58.8	36.9	50.6	53.1
SURE	53.8	53.8	33.8	59.4	69.4	54.0
Minimax	46.3	56.3	46.9	59.4	56.3	53.0

*(Note : single refers to the median value of details coefficients of the first level and adapt refers to median value for every detail coefficients at every level)

6. Discussions

From the results, it shows that a clean signal can be obtained with wavelet-based de-noising shrinkage method. The reconstruction algorithm recovers a close approximation of the original signal. It has shown that in Figure 2, the signal obtained when using universal and minimax threshold parameters visually appeared smoother compared to other thresholds. However, the smoothness of the signal does not influence the performance of the classification rate. The results in Table 1 show that the signals using heuristic threshold gives the best performance with average classification rate across all tasks of 82.3%, followed with SURE threshold which gives 75.4%. From Table 2, it shows that heuristic threshold again gives the best performance with average classification rate across all tasks of 64.6%. And heuristic threshold also gives the best performance when adapt median value is used as shown in Table 3. The best average classification rate (82.3%) from this study is achieved with heuristic threshold using default median value (=1). These threshold parameters can perform poorly if the coefficients are very sparse (most of the coefficients at a level are nearly zero), but heuristic threshold selection can adapt easily in this situation. It is sufficient to adopt the default median value (no rescaling) only as shown in all the results obtained. Overall average classification performance of the classifier had shown that threshold parameter using heuristic selection procedure outperforms the other thresholds.

7. Conclusions

Wavelet de-noising method can be adopted in improving the smoothness of EEG signals. Future work will investigate the de-noising procedure with different mother wavelets to improve the classification rate. These useful signals will be used as input for brain-computer interface specially developed for persons with severe physical disabilities.

Acknowledgements

This project is sponsored by Ministry of Higher Education, Malaysia.

References

1. Donoho D.L. *De-noising via Soft Thresholding*, IEEE Trans. Information Theory, May 1995, 41, pp. 613 – 627.
2. Donoho D.L. and Johnstone I.M. *Adapting to Unknown Smoothness via Wavelet Shrinkage*, J. Amer. Stat. Assoc., Dec. 1995, 90(432), pp. 1200-1224.
3. Taswell C. *The What, How, and Why of Wavelet Shrinkage Denoising*, IEEE Computing in Sci. and Eng., May/June 2000, pp. 12 – 19.
4. Misiti M. Misiti Y. Oppenheim G. and Poggi J.M. MATLAB Wavelet Toolbox User's Guide, 1998, The Mathworks, Inc.
5. Donoho D.L. and Johnstone I.M. *Minimax Estimation via Wavelet Shrinkage*, Ann. Statistic., 1993, 26(3), pp. 879-921.
6. Keirn Z.A. and Aunon J.I., *A New Mode of Communication Between Man and His Surroundings*, IEEE Trans. Biomed. Eng., 1990, 37(12), pp. 1209-1214.
7. Garett D., Peterson D.A., Anderson C.W. and Thaut M.H. *Comparison of Linear, Nonlinear, and Feature Selection Methods for EEG Signal Classification*, IEEE Trans. On Neural Sys. And Rehab. Eng., 2003, 11(2), pp.141-144.

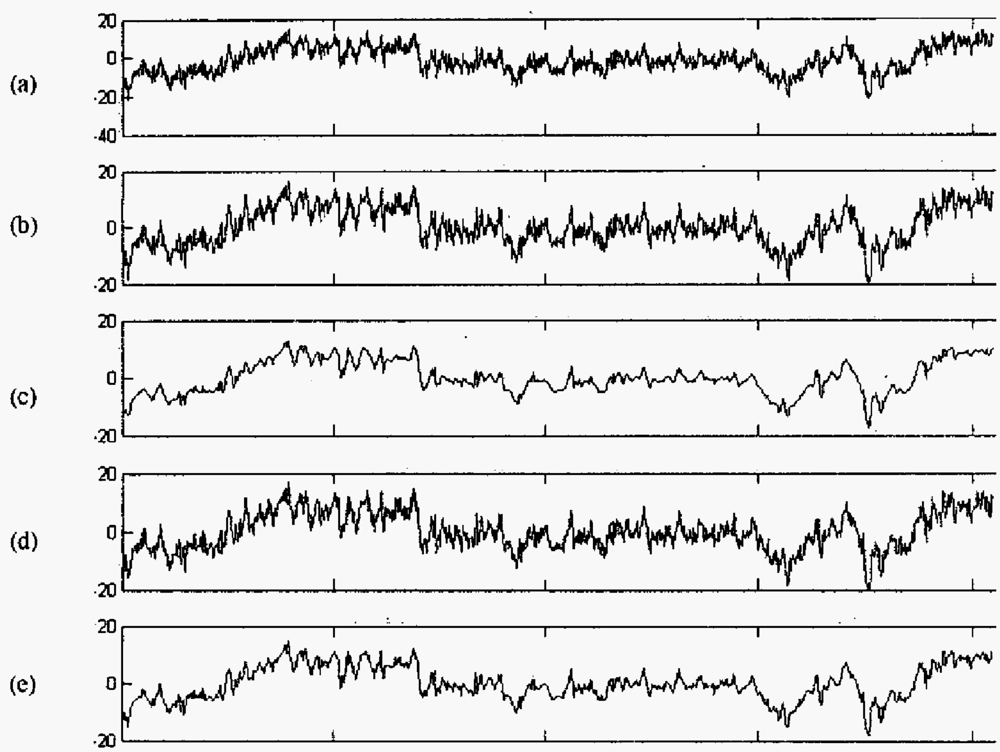

Figure 2 Original and De-noised Signals (a) Original Signal (b) SURE (c) Universal (d) Heuristic (e) Minimax